AF/2175
IFW

PATENT
Docket No. 0026-0001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of:	)	
	)	
Krishna Bharat et al.	)	
	)	
Serial No.: 09/729,240	)	Group Art Unit: 2175
	)	
Filed: December 5, 2000	)	Examiner: H. Mahmoudi
	)	
For: IDENTIFICATION OF SEMANTIC UNITS FROM WITHIN A SEARCH QUERY	)	

U.S. Patent and Trademark Office
220 20th Street South
Customer Window, Mail Stop Appeal Brief - Patents
Crystal Plaza Two, Lobby, Room 1B03
Arlington, Virginia 22202

APPEAL BRIEF

This Appeal Brief is submitted in response to the final Office Action, dated May 28, 2004, and in support of the Notice of Appeal, filed September 27, 2004 with a petition for a one month extension of time.

I. REAL PARTY IN INTEREST

The real party in interest in this appeal is Google Inc.

II. RELATED APPEALS, INTERFERENCES, AND JUDICIAL PROCEEDINGS

Appellants are unaware of any related appeals, interferences or judicial proceedings.

III. STATUS OF CLAIMS

Claims 1-41 are pending in this application. All the claims were finally rejected in the final Office Action.

Claims 1-3, 5-8, 10-15, 17-22, 24-27, 29-32, 34, and 36-41 stand rejected under U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,778,361 to Nanjo et al. ("Nanjo") in view of U.S. Patent No. 6,012,053 to Pant et al. ("Pant").

Claims 4, 9, 16, 23, 28, and 33 stand rejected under U.S.C. § 103(a) as being unpatentable over Nanjo and Pant, and further in view of U.S. Patent No. 6,385,602 to Tso et al. ("Tso").

Claim 35 stands rejected under U.S.C. § 103(a) as being unpatentable over Nanjo and Pant, and further in view of U.S. Patent No. 6,134,554 to Freimann et al. ("Freimann").

Claims 1-41 are the subject of the present appeal. These claims are reproduced in the Claim Appendix of this Appeal Brief.

IV. STATUS OF AMENDMENTS

No amendments have been filed subsequent to the final Office Action, dated May 28, 2004.

V. SUMMARY OF CLAIMED SUBJECT MATTER

In the paragraphs that follow, each of the independent claims and the claims reciting means-plus-function or step-plus-function language that is involved in this appeal will be recited followed in parenthesis by examples of where support can be found in the specification and drawings.

Claim 1 is directed to a method of identifying semantic units within a search query. The method includes identifying documents relating to the query (act 202; p. 9, second full paragraph) by comparing search terms in the query to an index of a corpus and generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words (p. 10, lines 6-21). The method further includes calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Further, the method includes selecting semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Claim 6 is directed to a method of locating documents in response to a search query. The method includes receiving the search query from a user (act 201) and generating a list of relevant documents based on search terms of the query (act 202; p. 9, second full paragraph). The method further includes identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents (act 203; p. 10, lines 1-5) and generating a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words (p. 10, lines 6-21). Still further, the method includes calculating, for each of the generated substrings, a value related to one or more documents in the subset of documents that contain the substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18) and selecting semantic units from the generated multiword substrings based on the calculated value (acts 208 and 209; p. 10, line 22 through p. 11, line 7). Additionally, the method includes refining the generated list of relevant documents based on the selected semantic units (p. 13, lines 10-13).

Claim 11 is directed to a system that includes a server (110) connected to a network

(101), the server receives search queries from users via the network. The server includes at least one processor (111) and a memory (112) operatively coupled to the processor. The memory stores program instructions that when executed by the processor, cause the processor to: identify a list of documents relating to the search query by matching individual search terms in the query to an index of a corpus (act 202; p. 9, second full paragraph) generate a plurality of multiword substrings from the query in which each of the substrings includes at least two words (p. 10; lines 6-21); calculate, for each of the generated substrings, a value relating to one or more documents of the identified list of documents that contain the generated substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18) and select semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Claim 18 is directed to a server (110) that includes a processor (111) and a memory (112) operatively coupled to the processor. The memory includes a ranking component (122) configured to return a list of documents ordered by relevance in response to a search query (act 202; p. 9 second full paragraph) and a semantic unit locator component (121) configured to locate semantic units, each having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component (acts 204-209 and 301-304; pages 9-13).

Claim 25 is directed to a computer-readable medium (112) storing instructions for causing at least one processor (111) to perform a method that identifies semantic units within a search query. The method includes identifying documents relating to the query by matching individual search terms in the query to an index of a corpus (act 202; p. 9, second full paragraph)

and forming a plurality of multiword substrings of the query in which each of the substrings includes at least two words (p. 10, lines 6-21). The method further includes calculating, for each of the substrings, a value relating to the portion of the identified documents that contain the substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Additionally, the method includes selecting semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Claim 30 is directed to a computer-readable medium (112) storing instructions for causing a processor (111) to perform a method. The method includes receiving a search query from a user (act 201) and generating a list of relevant documents based on individual search terms of the query (act 202; p. 9, second full paragraph). The method further includes identifying a subset of documents that are the most relevant documents from the list of relevant documents (act 203; p. 10, lines 1-5) and forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words (p. 10, lines 6-21). Additionally, the method includes calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18) and selecting semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7). Further, the method includes refining the generated list of relevant documents based on the selected semantic units (p. 13, lines 10-13).

Claim 36 is directed to an apparatus (201) for locating documents in response to a search query (act 201). The apparatus comprises means for receiving the search query from a user (110 and act 201) and means for generating a list of relevant documents based on individual search

terms of the query (122 and act 202; p. 9, second full paragraph). Further, the apparatus comprises means for identifying a subset of documents that are the most relevant documents from the list of relevant documents (122 and act 203; p. 10, lines 1-5), means for forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words (121 and p. 10, lines 6-21), and means for calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring (121 and acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Further, the apparatus includes means for selecting semantic units from the generated multiword substrings based on the calculated values (121 and acts 208 and 209; p. 10, line 22 through p. 11, line 7) and means for refining the generated list of relevant documents based on the selected semantic units (121 and p. 13, lines 10-13).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

A. Claims 1-3, 5-8, 10-15, 17-22, 24-27, 29-32, 34, and 36-41 stand rejected under U.S.C. § 103(a) as being obvious over Nanjo in view of Pant.

B. Claims 4, 9, 16, 23, 28, and 33 stand rejected under U.S.C. § 103(a) as being obvious over Nanjo, Pant, and further in view of Tso.

C. Claim 35 stands rejected under U.S.C. § 103(a) as being obvious over Nanjo, Pant, and further in view of Freimann.

VII. ARGUMENT

A. Rejection Under 35 U.S.C. § 103(a) over Nanjo and Pant

1. Claims 1, 2, 11, 14, 25, and 26

It is a cardinal tenant of patent law that to establish *prima facie* obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974). "All words in a claim must be considered in judging the patentability of that claim against the prior art." *In re Wilson*, 424 F.2d 1382, 1385, 165 USPQ 494, 496 (CCPA 1970). If an independent claim is nonobvious under 35 U.S.C. § 103, then any claim depending therefrom is nonobvious. *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988).

In rejecting representative claim 1, the Examiner contends that Nanjo discloses the first two elements recited in claim 1 but concedes that Nanjo does not disclose the last two elements of claim 1. (final Office Action, pages 2-3). The Examiner contends, however, that Pant cures the deficiencies of Nanjo and states that it would have been obvious to modify Nanjo in view of Pant to disclose the invention recited in claim 1. Appellants strongly disagree with the Examiner's assertions. In particular, as will be discussed below, Nanjo fails to disclose or suggest many of the elements recited in claim 1. Pant is similarly deficient and does not disclose or suggest the elements of claim 1 that the Examiner concedes are not disclosed by Nanjo. Thus, all of the claimed limitations are not taught or suggested and the rejection of this claim should be reversed.

Claim 1 is a method of identifying semantic units within a search query. Claim 1 includes identifying documents relating to the query by comparing search terms in the query to an

index of a corpus and generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words. Claim 1 further recites calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring and selecting semantic units from the generated multiword substrings based on the calculated values.

As an initial issue, Appellants and the Examiner have been unable to agree upon the meaning of the phrase "semantic unit." In previous responses, Appellants stated that the term "semantic unit," as defined by the Appellants' specification, refers to multiple terms that are considered to function as a "compound" that forms a single semantically meaningful unit. (See Spec., page 2). In response, the Examiner has refused to use this definition, stating "multiple terms that are considered to function as a 'compound' that forms a single semantically meaningful unit is not recited in the rejected claim." (Final Office Action, page 14). Instead, the Examiner appears to interpreting "semantic units" very broadly to cover virtually any text string(s). (See final Office Action, pages 2 and 14).

Appellants submit that the Examiner's definition of "semantic units" is overly broad and is inapposite to the plain meaning of the phrase. The Merriam-Webster Online dictionary, for instance, defines semantic as "of or relating to meaning in language." Thus, a multiword semantic unit, as recited in claim 1, refers to multiple terms related by meaning.

Additionally, Appellants note that an applicant is entitled to be his or her own lexicographer. *See In re Paulsen,* 30 F.3d 1475, 1480 31 USPQ2d 1671, 1674 (Fed. Cir. 1994). In this regard, Appellants' specification clearly defines and uses the term semantic unit consistent with the plain meaning of the phrase. At page 2, for instance, Appellants' specification defines

the term semantic unit (also called a compound in the specification) in the context of the example semantic unit "baldur's gate":

> Multiple search terms entered by a user are often more useful if considered by the search engine as a single compound unit. Assume that a user enters the search terms "baldur's gate download." The user intends for this query to return web pages that are relevant to the user's intention of downloading the computer game called "baldur's gate." Although "baldur's gate" includes two words, the two words together form a single semantically meaningful unit. If the search engine is able to recognize "baldur's gate" as a single semantic unit, called a compound herein, the search engine is more likely to return the web pages desired by the user.

Page 4 of Appellants' specification further elaborates on this definition:

> For example, the queries "country western mp3" and "leaving the old country western migration" both have the words "country" and "western" next to each other. Only for the first query, however, is "country western" a representative compound. Segmenting such queries correctly requires some understanding of the meaning of the query. In the second query, the compound "western migration" is more appropriate, although it occurs less frequently in general.

When using an appropriate definition for semantic unit, Appellants submit that Nanjo clearly does not disclose or suggest many of the features recited in claim 1.

Nanjo is directed to the indexing and searching of text in compound-word languages. (Nanjo, Title). Nanjo gives examples of such compound-word languages as Japanese, Chinese, Hebrew, and Arabic. (Nanjo, Abstract). A compound-word language is defined by Nanjo as "languages having words that are run together or lack intervening word separators, particularly Japanese, Chinese, or other Eastern languages." (Nanjo, col. 1, lines 10-14).

Nanjo discloses, among other things, the creation of a content index from a number of input documents. (Nanjo, col. 5, lines 22-42). In response to a user search query, the content index may be searched to obtain search results. (Id.). According to Nanjo, the content index may

include tokens in which a single string may yield multiple tokens. Nanjo discuses this feature in detail at column 6, lines 2-21:

> [T]he index is created by taking the collection of symbols forming the kanji character string, and creating a number of index terms each of a length the same as the step size, beginning with the first term in the string, and extending to the end of the kanji string, and thereafter progressively reducing the step size such that the last character in the kanji string is the last index term. In this manner, all kanji terms are taken in "chunks" of the step size or less, always beginning with one of the kanji symbols and always ending with a symbol at the end of a string of four or ending with the last symbol in the string.
>
> The reason for step indexing is to cause the system to treat every kanji symbol or character as the potential beginning of a word. Furthermore, a step size is utilized that is equal to or longer than most words in the language in question. For Japanese, a step size of four is believed to be optimal. The document is then indexed by all tokens produced by the step indexing method. For example, the string "abcdefg" yields the tokens "abcd" "bcde", "cdef", "defg", "efg", "fg", and "g".

Appellants submit that the tokens created using the step indexing method of Nanjo are not semantic units. In particular, Nanjo appears to simply take multi-character sub-strings of a larger string (i.e., sub-strings of a compound word). Nothing in Nanjo discloses or suggests that these sub-strings are formed from multiword phrases that form a semantically meaningful unit. In contrast, the strings processed by Nanjo appear to be single words or compound words. Appellants submit that a compound word such as "newspaper" is still a single word that in no way discloses or suggests the multiword semantic units recited in claim 1.

The Examiner points to column 5, lines 23-32 and column 8, lines 21-33 as disclosing identifying semantic units. (Final Office Action, page 2). Column 5, lines 23-28 of Nanjo states that the invention of Nanjo is particularly applicable in "compound-word languages such as Japanese or Chinese" and generally discusses searching and indexing the context-index. A

compound word is a single word and is thus not a semantic unit. Column 8, lines 21-33 of Nanjo states:

> The "Advanced Search" button 205, when depressed, generates an additional dialog (not illustrated), which allows the user to specify a more complex search criteria, for example a phrase search or a proximity search, or certain operators (e.g. Boolean AND, OR, NOT). Specifically, if the user wishes to specify a combination of text strings to search for, then the user uses the Advanced Search dialog to enter the text strings and the way in which the text strings should be combined. For example, the user could specify a search to find all documents containing the word "patent" or the word "application" or both words (sometimes denoted as "patent OR application").

This section of Nanjo relates to allowing the user to enter "advanced" search options for a search query, such as by manually specifying which proximity operators or Boolean operators are to be used with the search query terms. The Examiner apparently believes that allowing the user to specify a "combination of text strings," as disclosed in this section of Nanjo, discloses semantic units. (Final Office Action, page 2). Appellants respectfully disagree with the Examiner's interpretation of Nanjo. A "combination of text strings," as disclosed by Nanjo, in no way discloses or suggests a semantic unit. A search query entered by a user using Boolean operators, such as, for example, "country AND mp3" would be a "combination of text strings," as discussed by Nanjo, but is not a semantic unit.

In view of the above, Appellants submit that Nanjo is unrelated to the invention recited in claim 1, as Nanjo fails to disclose or suggest semantic units, as recited in this claim, much less the specific method recited by claim 1 for identifying semantic units. More specifically, regarding the recitations of claim 1, Nanjo does not identify semantic units within a search query by identifying documents relating to a search query by matching individual search terms in the

query to an index of a corpus and generating multiword substrings of the query in which each of the substrings includes at least two words. Further, Nanjo does not calculate values and select semantic units from the generated multiword substrings based on the calculated values, as recited in claim 1.

Appellants submit that Pant fails to cure the above-noted deficiencies of Nanjo. Pant is directed to a system for performing searches on a collection of information through which results from a search query are ranked according to user specified relevance factors that allow the user to control how the search results are presented. (Pant, Abstract). In the final Office Action, the Examiner alleged that Pant discloses "calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substrings (see column 7, lines 7-50), and selecting semantic units from the generated multiword substrings based on the calculated values (see column 10, lines 38-49, and see column 13, lines 9-25)." (Final Office Action, page 3). Appellants respectfully disagree with the Examiner's interpretation of the disclosure of Pant.

At column 7, lines 7-50, Pant discusses a number of relevance factors through which documents that contain at least some terms of a user search query can be ranked. For example, Pant states that "if a user enters a query that has six search terms, than documents which contain all six search terms are generally considered more relevant than documents which contain only five of the six search terms." (Pant, col. 7, lines 7-12). Another relevance factor mentioned by Pant "is the ordering of search terms in the document. That is, if the query terms appear in their given order in a document, than a relevance bonus may be applied to the document." (Pant, col. 7, lines 14-18). Column 10, lines 38-49 of Pant is similarly directed to weighting a document for

relevance based on the number of search terms that occur in the document.

Appellants submit that the relevance factors disclosed by Pant do not disclose or suggest the features of claim 1, including calculating values that correspond to comparisons between one or more identified documents and generated substrings, where the values are then used to select semantic units, as recited in claim 1. The sections of Pant cited by the Examiner are directed to the disclosure of relevance factors that are used to rank documents, not the calculation of values that are then used to select semantic units. Although Pant may generally calculate "values" based on the terms in a document and search queries, this does not disclose or suggest the features of claim 1. Further, Pant never mentions the word "semantic," much less anything related to identifying semantic units, as recited in claim 1.

For at least the foregoing reasons, Appellants submit that Nanjo and Pant, even if combined as the Examiner suggests, do not disclose or suggest many of the features recited in claim 1. In particular, neither Nanjo nor Pant disclose or suggest identifying semantic units within a search query, much less identifying semantic units in the specific manner recited in claim 1. Accordingly, it is respectfully submitted that claim 1 is not obvious under 35 U.S.C. § 103(a) in view of Nanjo and Pant. Reversal of the rejection of claims 1, 11, 14, 25, and 26 is respectfully requested.

2. Claims 6, 7, 30, 31, 36

Independent claim 6 is directed to a method of locating documents in response to a search query. Claim 6 recites a number of features similar to those recited in claim 1, including "generating a plurality of multiword substrings of the query" and "selecting semantic units from the generated multiword substrings based on the calculated values." For reasons similar to those

given above regarding claim 1, Appellants submit that Nanjo and Pant, either taken alone or in combination, do not disclose or suggest these features of claim 6.

Claim 6 includes additional features not disclosed or suggested by Nanjo and Pant, either alone or in combination. For instance, claim 6 recites "identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents." The Examiner contends that Nanjo, at column 17, lines 35-47, discloses this feature of claim 6. This section of Nanjo discloses:

> If at step 930 the type is not roman, the "no" branch is taken to step 943, and the inquiry is made whether the type is katakana. If so, the "yes"0 branch is taken to step 945 and the katakana string is formed into a string level search term. It should be understood at this juncture that for a katakana string, the entire index term list is searched. Any index term that includes the katakana string is considered a match, and all of its document associations are returned from the index. This is similar to a search in which the "*" wild card is added to the beginning and the end of the search string. Searching the entire index term list in this manner takes longer than a typical search of the index in which only exact matches are returned.

This section of Nanjo appears to generally discuss details by which search terms are matched to an index to return documents. In no way, however, could this section of Nanjo be considered to disclose to suggest identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents, as recited in claim 6. Nanjo appears to merely disclose identifying a set of documents based on search terms. Nanjo does not, however, then identify a subset of documents as recited in claim 6.

Claim 6 additionally recites "refining the generated list of relevant documents based on the selected semantic units." As previously mentioned, neither Nanjo nor Pant identify semantic units present in a search query, and accordingly, they could not possibly disclose refining a list of documents based on semantic units.

For at least the foregoing reasons, Appellants submit that Nanjo and Pant do not disclose or suggest each of the features recited in claim 6. Reversal of the rejection of claims 6, 7, 30, 31, and 36 is therefore respectfully requested.

3. Claim 18

Independent claim 18 recites a number of features, including a ranking component configured to return a list of documents ordered by relevance in response to a search query and a semantic unit component configured to locate semantic units, having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component. As previously discussed, neither Nanjo nor Pant discloses or suggests locating semantic units in search queries, much less locating semantic units based on a predetermined number of most relevant documents in a list of documents returned by the ranking component. Accordingly, Appellants submit that neither Nanjo nor Pant, either alone or in combination, could possibly disclose or suggest the semantic unit component recited in claim 18.

In rejecting claim 18, the Examiner points to various sections of Nanjo as disclosing the semantic unit component. In particular, the Examiner points to column 10, lines 1-4; column 8, line 33, and column 7, lines 15-18, as disclosing this feature of the invention. These sections of Nanjo are reproduced below:

> When a term is located in the directory structure 304, the leaf structure associated with the corresponding node is examined to retrieve the references to the documents that contain that indexing term.

(Nanjo, column 10, lines 1-4)

For example, the user could specify a search to find all documents containing the

> word "patent" or the word "application" or both words (sometimes denoted as "patent OR application").

(Nanjo, column 8, lines 29-33)

> FIG. 1 is an overview block diagram of the process used to generate a search result in accordance with the invention. A query 101 is generated by a program or by a user and sent as input to search system 102.

(Nanjo, column 7, lines 15-18). Appellants submit that these sections of Nanjo in no discloses or suggests the semantic unit locator component recited in claim 18. At best, they appear to relate to techniques for searching documents based on search queries applied to a content-index arranged in a directory structure.

For at least the foregoing reasons, Appellants submit that Nanjo and Pant do not disclose or suggest each of the features recited in claim 18. Accordingly, the rejection of claim 18 under 35 U.S.C. § 103(a) in view of Nanjo and Pant is improper and should be reversed.

4. Claims 3, 8, 15, 22, 27, and 32

Dependent claim 3 recites that the selection of the semantic units further includes selecting semantic units from the generated substrings that have calculated values above a predetermined threshold. The Examiner points to column 20, lines 41-50, of Nanjo and column 8, lines 57-62 of Pant as disclosing this feature. (final Office Action, page 4). This section of Nanjo corresponds to a feature of claim 1 of Nanjo that refers to "step indexing the symbols in the preliminary index term to create a plurality of index terms of a length equal to or less than a predetermined step size."

Appellants submit that the predetermined step size recited in this claim of Nanjo is not equivalent to, and does not suggest selecting semantic units based on calculated values above a

predetermined threshold. Nanjo merely creates a plurality of index terms by stepping through a preliminary index term using a predetermined step size. This does not disclose or suggest, however, comparing calculated values for substrings to a predetermined threshold, as required by claim 3.

The Examiner also relies on Pant as allegedly disclosing portions of claim 3. The section of Pant pointed to by the Examiner as disclosing the subject matter of claim 3 states:

> It should be understood that any other form of message that contains the search terms and relevance factors may be used to communicate them to the database query engine and that the invention is not limited to any particular form. The user also may specify a kind of search for which the weights associated with the relevance factors are predetermined.

(Pant, column 8, lines 57-62). Appellants submit that this section of Pant is in no way related to selecting semantic units from generated substrings that have calculated values above a predetermined threshold, as recited in claim 3.

Accordingly, for these reasons, the rejection of claims 3, 8, 15, 22, 27, and 32 should also be reversed.

5. Claims 5, 10, 17, 24, 29, and 34

Dependent claim 5 further defines the features of claim 1 and recites that "the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents."

The Examiner points to column 7, lines 23-30 and column 10, lines 36-49 of Pant as disclosing these features. (final Office Action, page 5). These sections of Pant, as previously discussed, relate to relevance factors that can be used to determine a relevance ranking for a

document. Calculating the relevance of a document according to a specified relevance factor, however, does not disclose or suggest the calculated values recited in claim 5, in which substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents. Pant completely fails to disclose or suggest this feature.

Accordingly, for these reasons, the rejection of claims 5, 10, 17, 24, 29, 34 should also be reversed.

6. Claim 21

Claim 21 depends from claim 18, and further recites that the semantic unit locator is further configured to, *inter alia*: calculate, for each generated substring, a value relating to the portion of the predetermined number of the most relevant documents that contain the substring; and locate the semantic units from the generated values. As discussed above, Nanjo may disclose generating multiple tokens from a single string, (see Nanjo, column 6, lines 2-21), Nanjo completely fails to disclose or suggest calculating the value recited in claim 21. Further, Nanjo does not even mention the concept of a semantic unit, as recited in claim 21, and therefore can not be said to "locate semantic units from the generated values," as is also recited in claim 21. Pant does not cure these deficiencies of Nanjo.

For at least the foregoing reasons, Appellants submit that Nanjo and Pant do not disclose or suggest each of the features recited in claim 21. Accordingly, the rejection of claim 21 under 35 U.S.C. § 103(a) in view of Nanjo and Pant is improper and should be reversed.

7. Claims 12, 13, 19, and 20

Claims 12, 13, 19, and 20 are dependent claims. Representative claim 12 recites that a

processor refines the identified list of documents based on the selected semantic units. The Examiner cites column 19, lines 15-25 of Nanjo as allegedly disclosing this feature. (final Office Action, page 8). Again, Appellants disagree with the Examiner's interpretation of Nanjo.

The cited section of Nanjo discloses:

> Also, according to this embodiment, it is preferable that a flag be included with each reference in the stored search result to indicate whether the reference was placed in the stored search result as a result of a direct search of the object as opposed to as a result of a search using the content-index. This flag is used for optimization purposes to avoid unnecessary searching of the object in the search result correction routines. One skilled in the art will recognize that the inclusion of such a flag is not necessary and that other implementations of preserving such information are possible.

(Nanjo, column 19, lines 15-25). This section of Nanjo relates to the use of an optional flag to indicate how a search result was obtained. Appellants are unsure how this disclosure of Nanjo even remotely relates to refining a list of documents based on semantic units. Accordingly, Appellants submit that neither Nanjo nor Pant, either alone or in combination, disclose or suggest the features recited in claim 12. Accordingly, the rejection of claims 12, 13, 19, and 20 under 35 U.S.C. § 103(a) should be reversed.

8. Claims 37-41

Claims 37-41 are dependent claims. Representative claim 37 recites that the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents. The Examiner alleges that this feature is disclosed by Pant at column 7, lines 7-50 and column 10, lines 38-49. (final Office Action, page 11). These cited sections of Pant were previously discussed, and

generally relate to different relevance factors that can be used to determine the relevance of a document to a search query. Although claim 37 does include the words "relevance" and "documents," claim 37 recites more than just determining document relevance. More specifically, claim 37 recites that the calculated "values are weighted . . . such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents." Pant completely fails to disclose or suggest weighting any values corresponding to the values recited in claim 37 or that the values are weighted in the manner recited in claim 37.

Accordingly, Appellants submit that neither Nanjo nor Pant, either alone or in combination, disclose or suggest the features recited in claim 37. Accordingly, the rejection of claims 37-41 under 35 U.S.C. § 103(a) should be reversed.

B. Rejection of Claims 4, 9, 16, 23, 28, and 33 Under 35 U.S.C. § 103(a) over Nanjo, Pant, and Further in View of Tso.

Claims 4, 9, 16, 23, 28, and 33 are dependent claims and claim 4 is representative of this group. In rejecting claim 4, the Examiner concedes that Nanjo does not disclose discarding generated substrings that overlap other ones of generated substrings with higher calculated values. (final Office Action, page 11). The Examiner contends, however, that Tso discloses this feature and that one of ordinary skill in the art would have found it obvious to modify Nanjo and Pant in view of Tso to obtain the features recited in these claims. (final Office Action, page 12).

Tso is directed to the presentation of search results in which the search results are dynamically categorized. According to Tso, search results are examined and one or more categories of results are established based upon attributes of the search results. (Tso, Abstract).

The categories of search results may then be presented using category indicators. (Id., Abstract).

The Examiner particularly points to column 5, lines 9-28 of Tso as disclosing the features recited in claims 4, 9, 16, 23, 28, and 33. (Office Action, page 12). Although these sections of Tso generally relate to discarding "search results" that are below a relevance threshold, the search results of Tso appear to be standard search results, such as a complete documents, and are not a generated substring, as required by claim 4. Accordingly, Appellants submit that Tso does not disclose or suggest the features recited in claim 4.

For at least these reasons, the rejections of claims 4, 9, 16, 23, 28, and 33 under 35 U.S.C. §103(a) should be reversed.

C. Rejection of Claim 35 Under 35 U.S.C. § 103(a) over Nanjo, Pant, and Further in View of Freimann.

Claim 35 is a dependent claim. In rejecting dependent claim 35, the Examiner relies on Freimann, in addition to Nanjo and Pant, for the disclosure of a computer-readable medium. Appellants have reviewed Freimann, and submit that the disclosure of Freimann does not cure the above-noted deficiencies in the disclosures of Nanjo and Pant. Accordingly, the rejection of claim 35 should also be reversed.

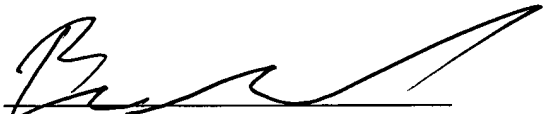
VIII. CONCLUSION

In view of the foregoing arguments, Appellants respectfully solicit the Honorable Board to reverse the Examiner's rejections of claims 1-41 under 35 U.S.C. § 103(a).

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.



Brian E. Ledell
Reg. No. 42,784

Date: November 24, 2004

11240 Waples Mill Road
Suite 300
Fairfax, Virginia 22030
(571) 432-0800

Customer No. 44989

CLAIM APPENDIX

1. A method of identifying semantic units within a search query comprising:

identifying documents relating to the query by comparing search terms in the query to an index of a corpus;

generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words;

calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring; and

selecting semantic units from the generated multiword substrings based on the calculated values.

2. The method of claim 1, wherein the identification of the documents further includes:

generating an initial list of relevant documents; and

selecting a predetermined number of most relevant ones of the documents in the initial list as the identified documents.

3. The method of claim 1, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

4. The method of claim 3, wherein the selection of the semantic units further includes:

discarding the generated substrings that overlap other ones of the generated substrings with higher calculated values.

5. The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

6. A method of locating documents in response to a search query, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on search terms of the query;

identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents;

generating a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

calculating, for each of the generated substrings, a value related to one or more documents in the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

7. The method of claim 6, wherein the identified subset includes a predetermined number of the most relevant ones of the documents in the list of relevant documents.

8. The method of claim 6, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

9. The method of claim 8, wherein the selection of the semantic units further includes:

discarding the generated substrings that overlap other ones of the generated substrings with higher calculated values.

10. The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

11. A system comprising:

a server connected to a network, the server receiving search queries from users via the

network, the server including:

at least one processor; and

a memory operatively coupled to the processor, the memory storing program instructions that when executed by the processor, cause the processor to: identify a list of documents relating to the search query by matching individual search terms in the query to an index of a corpus; generate a plurality of multiword substrings from the query in which each of the substrings includes at least two words; calculate, for each of the generated substrings, a value relating to one or more documents of the identified list of documents that contain the generated substring; and select semantic units from the generated multiword substrings based on the calculated values.

12. The system of claim 11, wherein the processor refines the identified list of documents based on the selected semantic units.

13. The system of claim 12, wherein the system transmits the refined list of documents to the user.

14. The system of claim 11, wherein the network is the Internet and the corpus is a collection of web documents.

15. The system of claim 11, wherein the memory includes instructions for causing the processor to:

select semantic units from the generated substrings that have calculated values above a predetermined threshold.

16. The system of claim 15, wherein the memory includes instructions for causing the processor to:

discard substrings that overlap other substrings with a higher calculated value.

17. The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

18. A server comprising:

a processor; and

a memory operatively coupled to the processor, the memory including:

a ranking component configured to return a list of documents ordered by relevance in response to a search query; and

a semantic unit locator component configured to locate semantic units, each having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component.

19. The server of claim 18, further including:

a search engine configured to refine the list of documents based on the located semantic units.

20. The server of claim 19, wherein the processor is configured to:

transmit the refined list of documents to a user that provided the query.

21. The server of claim 18, wherein the semantic unit locator is further configured to:

generate a plurality of substrings of the query;

calculate, for each generated substring, a value relating to the portion of the predetermined number of the most relevant documents that contain the substring; and

locate the semantic units from the generated values.

22. The server of claim 21, wherein the semantic unit locator is configured to locate semantic units from the generated substrings that have calculated values above a predetermined threshold.

23. The server of claim 22, wherein the semantic unit locator is configured to discard substrings that overlap other substrings with a higher calculated value.

24. The server of claim 21, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less

relevant documents.

25. A computer-readable medium storing instructions for causing at least one processor to perform a method that identifies semantic units within a search query, the method comprising:

identifying documents relating to the query by matching individual search terms in the query to an index of a corpus;

forming a plurality of multiword substrings of the query in which each of the substrings includes at least two words;

calculating, for each of the substrings, a value relating to the portion of the identified documents that contain the substring; and

selecting semantic units from the generated multiword substrings based on the calculated values.

26. The computer-readable medium of claim 25, wherein the identification of the set of documents further includes:

generating an initial list of relevant documents; and

selecting a predetermined number of the most relevant documents in the initial list to include in the identified documents.

27. The computer-readable medium of claim 25, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

28. The computer-readable medium of claim 27, wherein the selection of the semantic units further includes:

discarding substrings that overlap other substrings with a higher calculated value.

29. The computer-readable medium of claim 27, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

30. A computer-readable medium storing instructions for causing a processor to perform a method, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on individual search terms of the query;

identifying a subset of documents that are the most relevant documents from the list of relevant documents;

forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

31. The computer-readable medium of claim 30, wherein the identified subset includes a predetermined number of the most relevant documents from the list of relevant documents.

32. The computer-readable medium of claim 30, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

33. The computer-readable medium of claim 32, wherein the selection of the semantic units further includes:

discarding substrings that overlap other substrings with a higher calculated value.

34. The computer-readable medium of claim 30, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

35. The computer-readable medium of claim 30, wherein the computer-readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data signal embodied in a carrier wave.

36. An apparatus for locating documents in response to a search query, comprising:

means for receiving the search query from a user;

means for generating a list of relevant documents based on individual search terms of the query;

means for identifying a subset of documents that are the most relevant documents from the list of relevant documents;

means for forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

means for calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

means for selecting semantic units from the generated multiword substrings based on the calculated values; and

means for refining the generated list of relevant documents based on the selected semantic units.

37. The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the

substring in a less relevant one of the documents.

38. The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

39. The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

40. The computer-readable medium of claim 27, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

41. The computer-readable medium of claim 30, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	)	
	)	
Krishna Bharat et al.	)	Group Art Unit: 2175
	)	
Application No.: 09/729,240	)	Examiner: Hassan Mahmoudi
	)	
Filed: December 5, 2000	)	
	)	
For: IDENTIFICATION OF SEMANTIC UNITS FROM WITHIN A SEARCH QUERY	)	

TRANSMITTAL FOR APPEAL BRIEF

U.S. Patent and Trademark Office
220 20th Street S.
Customer Window, Mail Stop Appeal Brief - Patents
Crystal Plaza Two, Lobby, Room 1B03
Arlington, Virginia 22202

Sir:

Transmitted herewith is an Appeal Brief in support of the Notice of Appeal filed September 27, 2004.

Enclosed is a check for ☐ $170.00 ☒ $340.00 to cover the Government fee.

The Commissioner is hereby authorized to charge any appropriate fees under 37 C.F.R. §§ 1.16, 1.17, 1.20(d) and 1.21 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 50-1070.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.

By: _____
Brian E. Ledell
Reg. No. 42,784

11240 Waples Mill Road, Suite 300
Fairfax, Virginia 22030
(571) 432-0800
Customer Number: 44989

Date: November 24, 2004